Machine Reading

The Universal Text to Knowledge Engine

Distribution Statement "A" – Case # 12224

Building the Universal Text to Knowledge Engine

Build a universal engine that **captures knowledge** from naturally occurring **text** and **transforms** it into the **formal representations** used by AI's **reasoning systems**.

Distribution Statement "A" – Case # 12224

KEY PROBLEMS
1. Reading is inherently ambiguous at many linguistic and logical levels
2. Reading requires many implicit inferences

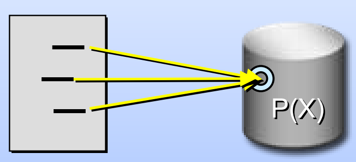
SOLUTION PARAMETERS

- Targeted to a pre-specified ontology
- Employs:
  - small amount of annotated text,
  - human tweaking, and
  - large amount of un annotated text
- General purpose & high performance by learning/bootstrapping/crafting ontology-specific reading systems.
- Not just a translation, but must bridge the mismatched assumptions in both corpora. (Can't do that in general, so it is done on an inference chaining by inference chain basis.)

PROMISING ENABLERS FOR THIS CHALLENGE

Consistency Trumps Ambiguity

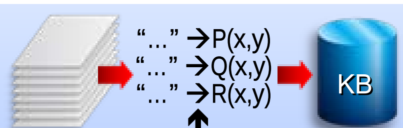
- Expect consistent subjects
- Require consistent theories



Evidence from Möbius

Learned Reading Patterns

- Unifying syntactic and semantic patterns
- Learn: manual encoding too expensive



[requires understanding linguistic patterns and semantic jumps]

Evidence from the field

Leverage Usage Context

Provide scaffolds for output of reading systems: natural and powerful.

Counter-intuitively, the *more* levels of comprehension tackled, the *easier* ambiguity resolution (reading) becomes.

Bringing power of formal reasoning to text, where most human knowledge is encoded

Why is reading so hard anyway?

ANSWER #1: **Because reading is INHERENTLY AMBIGUOUS at all levels.**

- **Pronoun reference ambiguity**
- **Parse structure ambiguity**
- **Word sense ambiguity**
- **Conceptual mapping ambiguity**

`“John chose the gun under {duress | warranty}.”`

There is no single correct syntactic parsing rule

ANSWER #2: **Because reading depends on millions of “immediate” inferences that are so automatic we forget that we are making them.**

Frame Axioms **(immediate inferences about what stays true)**

First read:	...then later read:	...new expectation
“Dan is cold”	**“Dan is hot”**	**(Dan *IS NO LONGER* cold)**
“Dan is friends w Bill”	**“Dan is friends w Jack”**	**(Dan *IS STILL* friends w Bill)**
“Dan is in the kitchen”	**“Dan is in the bedroom”**	**(Dan *IS NO LONGER* in kitchen)**
	“Dan is in New York”	**(Dan *IS STILL* in kitchen)**

← Just this one type of immediate inference requires a mini-theory for *EACH NEW CONCEPT*

Next slides enumerate key enablers

Two key problems... *Both* require massive knowledge to solve

Consistency Trumps Ambiguity

Ambiguity can be resolved by requiring...

Passage Consistency

Consistent interpretation over a text passage

Anaphora Resolution

"Dave collided with Jim. He [? Jim / Dave] was unharmed."

"Jim, on the other hand, suffered a concussion."

Passage consistency resolves pronoun ambiguity.

Word Sense Disambiguation

"You must adjust your attitude [(From WordNet) Mental state / Relative craft orientation / ⋮]."

Consistency disambiguates word sense choices.

"Your sullenness won't do." ➔ (*Mental state*)

"You're going to stall." ➔ (*Craft orientation*)

Theory Consistency

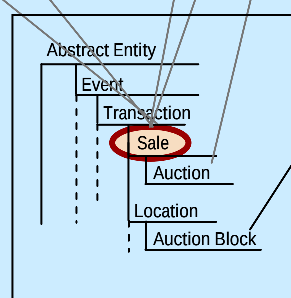
Consistency in the KB output from reading

"An auction is a process of buying and selling goods by offering them up for bid, taking bids, and then selling the item to the winning bidder."

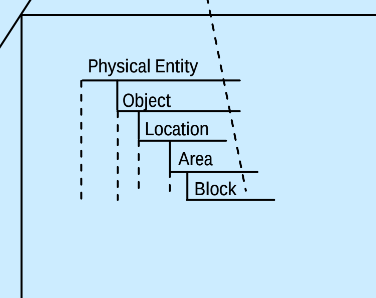
"Please join us for our first live sale of 2008 at Fredericksburg Auto Auction. Vehicles may be withdrawn from sale for continued government use at any time."

"Come on out to the Dakota auction, there are quite a few cars on the block today."

Output theories provide usage contexts that constrain possible interpretations of texts.



Theory of Commerce



Theory of Urban Geography

Supporting evidence from Semantic Elaboration process in Möbius

Distribution Statement "A" – Case # 12224

Text and Knowledge Contexts Reduce Ambiguity

Learned Reading Patterns

Resolve Ambiguity and Support Variation

Millions of content-specific patterns are key for
- **Resolving many levels of ambiguity**
- **Supporting the "immediate" inferences required by reading**

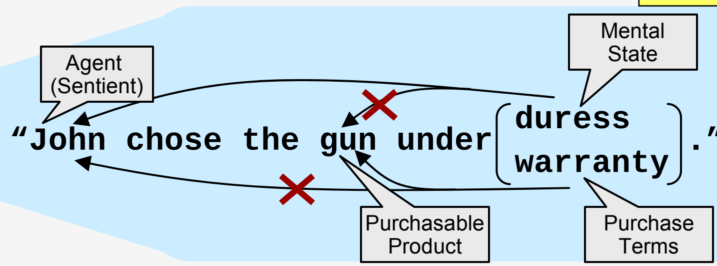
Learning is the *only* solution for acquiring these millions of patterns

RESULT: Generality – **learning to read is automated across all sub-disciplines.**
Performance – **patterns & inferences specific to concepts in each sub-discipline.**

Learned Patterns to Resolve Ambiguity

Must learn millions of these patterns

- **Pronoun Ambiguity**
- **Parse Ambiguity**
- **Word Sense Ambiguity**



***Pattern Used*: State of mind must be appropriate property of agent.**

***Pattern Used*: Only products and services have warranties/guarantees**

Learned Patterns to Support Inference

Take a binary relation, e.g., X solves Y.

- Collect examples of "X solves...", and "...solves Y"
- Find other relations with a ***similar distribution***

The following can solve:
people (116)
computers (37)
…

The following can rectify:
people (97)
groups (23)
…

The following can be solved:
problems (145)
issues (43)
…

The following can be rectified:
issues (56)
alternating current (23)
…

Learned Inference Patterns
(Statements learned that imply "*X solves Y*")

1. X tackles Y	7. X eases Y
2. X resolves Y	8. Y is solved by X
3. X finds a solution to Y	9. X alleviates Y
4. X rectifies Y	10. X corrects Y
5. X tries to solve Y	11. X is a solution to Y
6. Y is resolved by X	12. Y is blamed for X

(incorrect rule highlighted)

Today, this can be done at a low level.* However, this program requires learning such patterns at all levels.

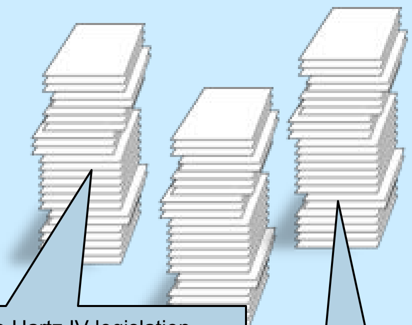
*Dekang Lin and Patrick Pantel. 2001. **Discovery of Inference Rules for Question Answering.** *Natural Language Engineering* 7(4)[343-360].

Key is to <u>learn</u> the millions of required patterns – manual coding is impractical.

Leverage Usage Context & Prior Knowledge

Specified by DARPA

Corpus



The Hartz IV legislation effectively abolished the current unemployment assistance by replacing it with a new so-called 'unemployment benefit II'..

After Social Democratic chairman Beck told state SPD party organizations they could forge their own alliances with the Left Party, his leadership has been criticized by more centrist SPD members.

Value of giving a target:

- Leverage Context
- Connect with all of AI
- Precise Go/No-Gos

Q/A Context

Political Analysis

VotedFor(?agent , ?bill)

CorePosition(?agent, ?pos)

Supports(?agent , ?position)

Agent

Political Group

Teamsters Union

Citizen

Legislator

"**Janik** defied his party and voted in favor of the Unemployment Benefit II bill."

"The Christian Democrats were unanimous in their support of the Hartz IV bill when it was voted on in the Bundesrat."

"The current Chancellor, Angela Merkel, began her tenure in November 2005 and is the first woman to hold this position."

Volker Kauder is a German CDU politician. He is Chairman of the CDU/CSU parliamentary group in the Bundestag.

Optionally supplied knowledge

Prior knowledge

[Seed Ontology]

P→Q [Patterns]

P(x) [Facts]

Queries

Q: Which legislators have cast votes on unemployment bills that differ from their stated positions?

Find All person **where**

VotedFor(person, bill) **and**
Supports(person , position1) **and**
Supports(bill , position2) **and**
position1 ≠ position2

A: "Heiner Janik, Susanna Tausendfreund,… "

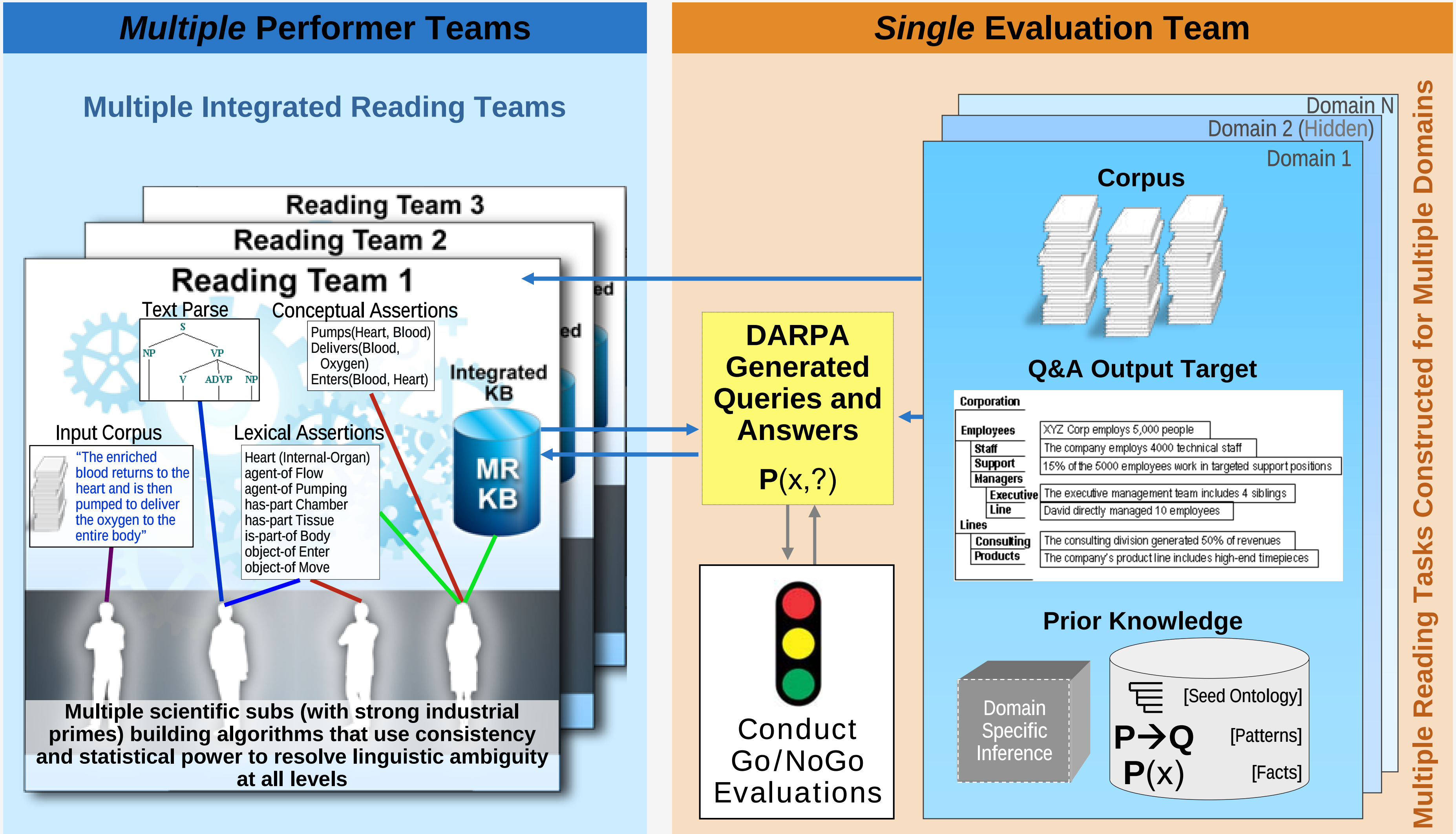
Q: Which groups could form a coalition without disagreeing on their core issues?

Find largest set { groupi } **where**
for all issue, group1, group2
if CorePosition(group1, position) **then**
Supports(group1 , position) =
Supports(group2 , position)

A: "SDP & Green,"
"PBC & CDU,"
"LINKE, GRAUE & SDP"…

Highest performance by employing all available knowledge

Evaluation and Performer Teams



Disparate Approaches; Universality enforced by multiple & hidden domains

Distribution Statement "A" – Case # 12224

Measuring Success

Head to Head: Man vs Machine

Performance Ratio (PR)

PR = F machine / F human

Reader's answer

✓ ✓ ✓ ✓ ✓ ✓ ✓ x x x x

Gold standard answers

false negatives (fn)	true positives (tp)	false positives (fp)

Correctness: fraction of retrieved answers that are correct.

Completeness: fraction of correct answers that are retrieved.

F MEASURE

$$F = \frac{2PR}{P + R}$$

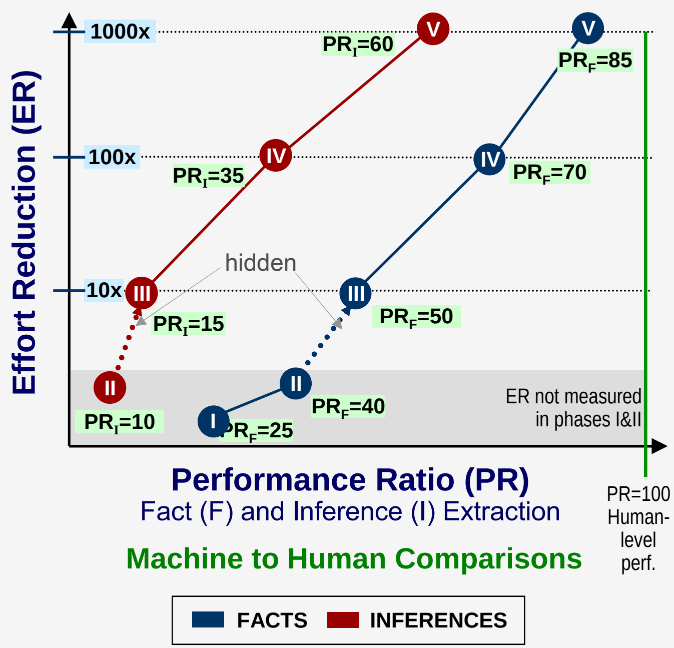
CORRECTNESS Precision (P)

$$P = \frac{tp}{tp + fp}$$

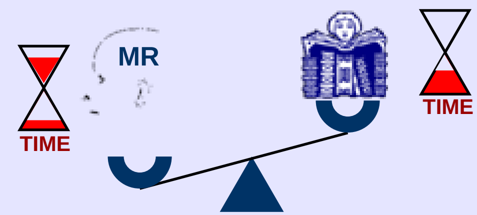
COMPLETENESS Recall (R)

$$R = \frac{tp}{tp + fn}$$

Go/No-Go Targets by Phase



Effort Reduction (ER)



Time to configure MR system VS Time to manually encode the KB.

EFFORT REDUCTION (ER)

CONSERVATIVE estimate of time for knowledge expert to manually code rules.

$$ER = \frac{\text{Manual coding time}}{\text{MR specialization time}}$$

Time to work with the MR system by knowledge engineer and adjust it for target.

Comparing automated reading to the manual alternative

Go/No-Go Criteria

GO/NO-GO THRESHOLDS	I	II	III	IV	V
Text Readability	100% of Humans				
Performance Ratio$_{Facts}$		40	50	70	85
Performance Ratio$_{Inference}$		10	15	35	60
Effort Reduction			10 x	100x	**1000x**

Distribution Statement "A" – Case # 12224

PHASE THRUSTS

Phase I:
End-To-End Capability

- **End to end reading system** TEXT → KNOWLEDGE converter
- **Multiple Targets** (2)
- **Ambiguity Resolution** (statistically) driven by internal consistency

Phase II:
Learn Reading Patterns

Auto-Learning Ambiguity Resolution
parse structure • part of speech • word sense • pronoun reference • mapping onto target • ...
Auto-Learning Inference Patterns
for surface form variation

Phase III:
Generality / Effort Reduction

- Hidden Targets
- Diversity of total targets (6)
- 10X **human effort reduction**
- **Military Transition Program:** (domain specific semantic search over heterogeneous MIL KBs)

Phase IV:
Inferential Complexity / Model Extrapolation

Auto-Learning inference patterns
- **"Immediate" Inference Rules** frame axioms • pragmatics • one-step interferences • ...
- Learning & **grouping knowledge** into inferentially deep logical theories
- Rudimentary **modal reasoning**

Phase V:
Scaling

Extend and adapt algorithms to operate at web scale

Note: GO/NO-GO requires average over "hidden" domains AND average over all prior domains to both meet all targets.

COMPLEXITY DIMENSIONS

	I	II	III	IV	V
Shifts in Complexity ➔	Consistency	Translational	Generative	Inferential	Hidden term discovery
Number of test domains	2	4	6	8	10
number of hidden domains			2	2	2
Largest Corpus Size	Homogenous texts	Heterogeneous texts 1,000 pages	2K pages	10K pages	Web Scale (~1M pages)
Linguistic constructs	Static descriptive K	Events	Temporal Inference	Modal Reasoning	Temporal Processes

Common metric + increasing number of tests + increasing scores + qualitative shifts in every phase.

DARPA: Solving Key AI Problem

Ambitiously aims where none have before:

The Universal Reading Machine

that maps any natural text into formal knowledge

Mountains of Knowledge as Text

Islands of formal AI knowledge

Distribution Statement "A" – Case # 11668